Data Science Self-Study Resources

Galvanize Inc.

Contents

Working through these questions is a great way to gauge your level of understanding in Data Science.

NOTE: This text can be shared publicly!

Python Resources

Python is an extremely popular and powerful programming language with tools to work in nearly every application area. In particular, Python (and specifically the `anaconda` environment that supports it) is the industry standard for machine learning. As such, comfort with python is a must for any aspiring data scientist.

In the Galvanize Data Science Immersive we teach data science which students build on top of their python expertise—we only briefly review python topics. We therefore require that students pass a python coding interview.

The following are various resources targeted to different levels of applicants:

Beginner (Little to no Experience)

If you've not coded much (or at all) before, "Learn Python the Hard Way" is a good way to understand and learn the basics of Python (and programming in general).

Taking the Python Fundamentals workshop offered at Galvanize is also a really great way to ramp up on python! Even better, if you take the workshop and are accepted into the Galvanize Data Science program, we will credit the amount you paid for the workshop to your tuition balance.

In addition, you can check out:

- Learn Python the Hard Way (http://bit.ly/2cmDtZt)
  - Specifically Exercises 0-39
- Codecademy's Python Course (http://bit.ly/1GO0Fx1)
- Galvanize Part-Time Course: Python Fundamentals (http://bit.ly/2jaJ7S2)

Novice (Some Experience Coding)

If you already understand the basics of programming, you may prefer the slightly more advanced treatment of Python in the Google Python course (http://bit.ly/1bmXyG5).

Intermediate (Programming Experience Needs a Refresher)

If you're already comfortable coding, you may just want to jump straight to the Python documentation. The tutorial available from the python site (http://bit.ly/1LBc6Ua) is a comprehensive and excellent resource to quickly transition from your current programming language(s) into Python.

Additionally, there is a higher-level Introduction to Computer Science and Programming using Python (http://bit.ly/2fLUXPC) course offered through edX that provides insights into more complex subjects such as sorting and runtime analysis. While this is more advanced than admittance into the Data Science Immersive requires, it can serve to take your Python skills to the next level.

Remember: To pass the python interview one must be able to comfortably solve most CodeWars (http://bit.ly/2iWF0LL) Level 6-7 problems in about 30 minutes!

Statistics Resources

In the Galvanize Data Science Immersive, we review probability, statistics, and classical regression but then build on these topics as we learn advanced and modern machine learning methodology. It is therefore extremely important that our students be quantitatively and analytically literate with strong comfort in baseline probability and statistical theory. We evaluate technical analytical skills though a Statistics/Machine Learning interview.

To pass the interview one should be able to. . .

1. Comfortably solve probability problems
2. Leverage statistical distributions to solve real world problems
3. Have a solid understanding of chapters 1-4 of an Introduction to Statistical Learning (http://bit.ly/1Wma81Z)

Novice (Little to No Experience)

If you've never seen statistics in your life (or can't remember anything about it) then an online "Introduction to Statistics" course would would be a good start. Khan Academy provides some good "entry level statistics videos" and a good introductory course is available on Udacity.

- Udacity's Intro to Statistics (http://bit.ly/1SnQHUx)
- Khan Academy: Probability (http://bit.ly/2j0KpBR)
- Khan Academy: Random Variables (http://bit.ly/1Q8I7FX)
- JB Statistics (http://bit.ly/2iWzSap)

Intermediate (Some Experience)

If any topics in the 14 sections covered in this Probability Review (http://bit.ly/2i6OLb4) remain unclear or fuzzy after you've gone over it you should search out some additional learning material online.

Going Further

Machine Learning

In addition to the probability and statistics content on the Stats/ML interview, key machine learning concepts will be discussed. A good overview of ML is available through the Introduction to Machine Learning course (http://bit.ly/2bp2YJX) videos (and the associated pdf textbook).

If you familiarize yourself with the content of that course you will come into Galvanize ready to maximize your learning experience.

SQL Resources

Structured Query Language (SQL) is a building block of many things we do in the Data Science world. If you are familiar with programming or data analysis, learning SQL won't take very long.

- SQLZoo (http://bit.ly/1RAqIc6) (Sections 0 to 7)
- Codacademy SQL Course (http://bit.ly/1QjodIp) (Chapters 1-4—not the pro version)

Self Assessment Questions

Have you been thinking about applying to Galvanize but are unsure whether you are ready for the Data Science Immersive? How do you know you have enough of a technical background to be successful in our Data Science Immersive course?

If you can answer most of the following questions, we want to see your application!

- Can you solve most Level 6-7 CodeWars challenges in Python in about 30 minutes?

- Can you explain the answers to the following scenarios?
  - If two events are independent, what is the probability that they both occur?
  - If two events are mutually exclusive, what is the probability that one or the other occurs?
  - What is the purpose of linear regression? How is the "best fit" line calculated?
  - What is the purpose of logistic regression? How is logistic regression similar to linear regression? How does it differ?
  - Have you ever performed a hypothesis test?

Comprehensive Introduction

Our Data Science Primer (http://bit.ly/1TnYbZu) is the most comprehensive resource and is an excellent preparation for the program. It is a substantial packet that may take some time however we would still recommend candidates complete this.

Admissions Process Description

Technical Assessment sent via email (3 questions, 4-hour time limit)

1. Python: The Python question won't require any special libraries. It will just be a basic programming task testing your ability to structure a program.
2. SQL: The SQL material is all about querying (select statements). We will provide an outline of a table, and then you will have to write a query to pull information from the table.
3. Statistics: The stats question will mainly concern hypothesis testing.

Python Interview via Google Hangout

This is a technical interview where you will solve a challenge problem in Python. It is conducted by a Data Science alum or instructor, via Google Hangout. You will use a collaborative code editor so the instructor can observe you as you work through the problem(s). The candidate should be familiar with all of the basic data structures in python, and basic control flow/logical statements. The candidate should be capable of writing simple functions in Python.

Statistics Interview via Google Hangout

The interview is conducted by an instructor or alum, via Google Hangout. There will be no software or programming; questions will be shared over chat and then worked through with the interviewer. You should be familiar with college level statistics and probability concepts. You may be asked about probability distributions, hypothesis testing, confidence intervals, multiple linear regression, logistic regression, and statistical modeling, as well as basic visual methods of model evaluation.

After the final interview, an Admissions decision will then be made within 1-2 business days.

NOTE: As tempting as it is to ask the interviewer whether you passed, they have been instructed to leave that final judgement to the admissions officer. Feel free, however, to ask your interviewer where you could improve and bring up any questions you have about the immersive!